International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques

[1]S.Hemasri,[2]V.Vamsikrishna,[3]V.Spandana,[4]Y.Sabareesh

UG students, Dept of CSE, Kallam Haranadhareddy Institute of Technology, AP, India.

Abstract

In both regulatory bodies in charge of transportation and regular people, road traffic safety is a top priority. Roadway traffic data must be carefully analyses in order to identify factors that are strongly correlated with safe driving advice deadly mishaps.In an effort to solve this issue, we use statistics analysis and data mining algorithms on the FARS Fatal Accident data set. The connection investigating the relationship between the fatality rate and other factors including driving under the influence, weather, road conditions, and surface conditions. The Apriori method is used to find association rules, the Naive Bayes classifier is used to create classification models, and the K-means clustering technique is used to create clusters.Based on statistics, association rules, a classification model, and the discovered clusters, certain driving safety recommendations are made.

Keyword - Roadway fatal accidents, association, classification, clustering.

1.INTRODUCTION

Many nations have felt the effects of globalisation. Travel and transportation have expanded as a result of a sharp growth in economic activity and consumption. Every day, there are numerous vehicles on the road, and incidents involving traffic can occur anywhere, anytime. Road accidents are caused by the growth of automobiles and traffic. Accidents that result in fatalities result in deaths of people. We all wish to keep safe and prevent accidents as humans. Given the significance of road safety, it is crucial to pinpoint the root causes of accidents in order to lower the number of incidents. It is challenging to examine the limitations leading the traffic accidents because of the exponential growth in accident data. The obtained data set must therefore be mined for common patterns that contribute to traffic accidents. There are far too many fatalities caused by road accidents each year. To prevent traffic accidents, the underlying causes must be identified. In order to uncover potential hidden patterns in gathered datasets depicting actual traffic incidents, data mining techniques must Links and connections between different elements impacting fatal traffic accidents. The outcomes of a data mining approach can aid in understanding the most important variables or often occurring patterns. The most hazardous roads in terms of traffic accidents are identified by the created pattern, and the appropriate precautions can be taken to avoid accidents on such roads.

According to figures from the World Health Organization, 1.3 million individuals worldwide lose their lives in traffic accidents every year, raising concerns about global road safety. The problem is considerably more concerning for low- and middle-income nations because, while having about 60% of the world's vehicles, these nations account for 93% of all traffic fatalities worldwide. Since that India has one of the greatest road networks in the world, the issue of road safety is even more crucial. The issue has been made worse by the unheard-of rate of motorization and the expanding urbanization brought on by the rapid rate of economic expansion.The population is a persistent issue in our nation.

The population of automobiles is growing together with the population of humans. Both the accident rate and the level of pollutants have increased as a result. Accidents have caused us the most trouble. Because of their concern for accidents, people hesitate before leaving their homes. Most often, we place the responsibility on the number of wheels on the car, but occasionally, we fail to acknowledge our own culpability. Sometimes, breaking driving laws might have even more catastrophic repercussions. We don't want people wearing helmets only to avoid paying a $100 fine; we want them worn for safety reasons.The most common things individuals do these days include texting or using their phone while driving, driving while intoxicated or recklessly, and going above the speed limit only to show off how fast their car is and how stylish they are. We won't always be given a second opportunity in life. This debate goes on forever. But, there is still a chance that technology will one day help us solve this. By knowing which area of the city is most prone to accidents, we can still save our people.

2. **LITERATURE SURIVEY**

Mussone (1999) used neural networks to analyse vehicle accidents that occurred at intersections in Milan, Italy. These authors used feed-forward multi layer perception (MLP) with BP learning. The model had 10 input nodes for eight variables: day/night, traffic flows in the intersection, number of virtual and real conflict points, intersection type, accident type, road surface condition, and weather condition. The output node (‘accident index’) was calculated as the ratio between the number of accidents at a given intersection and at the most dangerous intersection. Results showed that the highest accident index for the running over of pedestrians occurred at non-signalized intersections at night time.

Ossenbruggen, Pendharkar (2001) used a logistic regression model to identify the prediction factors of crashes and crash-related injuries, using models to perform a risk assessment of a given region. These models included attributes describing a site by its land use activity, roadside design, use of traffic control devices, and traffic exposure. Their study illustrated that village sites were less hazardous than residential or shopping sites. Abdalla et al. (1997) also studied the relationship between casualty frequency and the distance of an accident from residential zones. Not surprisingly, casualty frequencies were higher in accidents that occurred nearer to residentical zones, possibly due to higher exposure. The casualty rates among residents from relatively deprived areas were significantly higher than those from relatively affluent areas.

Sohn and Hyungwon (2001) conducted research on pattern recognition in the framework of RTA severity in Korea. They observed that an accurately estimated classification model for several RTA severity types as a function of related factors provided crucial information for accident prevention. Their research used three data mining techniques, neural network, logistic regression, and decision tree, to select a set of influential factors and to construct classification models for accident severity. Their three approaches were then compared in terms of classification accuracy. They 4found that accuracy did not differ significantly for each model, and that the protective device was the most important factor in the accident severity variation.

Ng, Hung and Wong (2002) used a combination of cluster analysis, regression analysis, and geographical information system (GIS) techniques to group homogeneous accident data, estimate the number of traffic accidents, and assess RTA risk in Hong Kong. Their resulting algorithm displayed improved accident risk estimation compared to estimates based on historical accident records alone. The algorithm was more efficient, especially for fatality and pedestrian-related accident analyses.

The authors claimed that the proposed algorithm could be used to help authorities effectively identify areas with high accident risk, and serve as a reference for town planners considering road safety.

Eric M Ossiander and Peter Cummings (2002) Investigated the effect on road speed on accident in the state of Washington. Some researches claim that those states which increased speed limit from 55mph to 65mph after 1974 had the fatality rate go up by 27% compared to increase in 10% in the states that did not increase the speed limit. It is claimed that as the effect of change in maximum speed is varies between urban and rural areas. After 1987 accident in rural areas increased while urban areas stayed relatively constant but clash rate in urban intersection is twice as high as in rural intersection. Accident is dependent on area (urban/rural), type of street (intersection, highway).

3. PROBLEM STATEMENT

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident. As human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion. In this project we apply statistics analysis and data mining algorithms on the FARS Fatal Accident dataset as an attempt to address this problem.

4. EXISTING SYSTEM

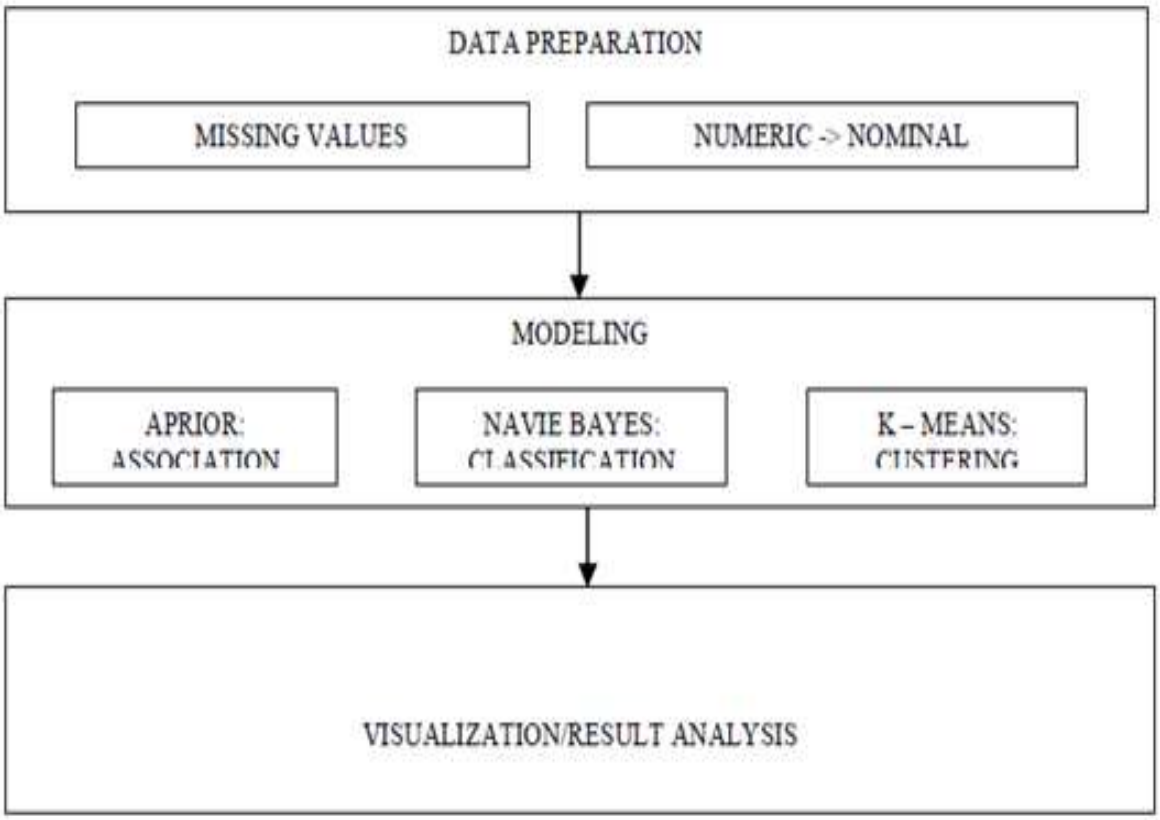
Road accident is one of the major causes of unnatural deaths , disability and property damage. Generally , when an accident took place , it is recorded by the concerned police officer of that region's police station .Police stations only cover the accidents which have happened in their territories. Poisson models and negative binomial models have been used extensively to identify the relationship between traffic accidents and the causative factors.By using these techniques in road accident data analysis may lead to some problems if the accident data have higher dimensions.

5. PROPOSED SYSTEM

In this we will apply statistical analysis and data mining algorithms on the FARS Fatal Accident data set as an attempt to address this problem. The relationship between fatal rate and other attributes including collision manner, weather condition, surface condition, light condition, and drunk driver.Association rules will be discovered by Apriori algorithm, classification model will be built by Naive Bayes classifier, and clusters are formed by simple K-means clustering algorithm.Certain safety driving suggestions are made based on statistics, association rules, classification model, and clusters obtained.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

6. BLOCK DIAGRAM



7. METHODOLOGY

7 .1 Data Cleaning

The raw form of statements, numbers, and qualitative phrases continues to be the primary data gathered from the web sources. Errors, omissions, and inconsistent data are present in the raw data. After carefully reviewing the filled surveys, it needs corrections. The processing of primary data involves the following processes. For equivalent details of individual responses, a sizable volume of raw data gathered from field surveys needs to be organized. Data Pre-processing is a method used to turn the unclean raw data into a clean data set. In other words, when data are received from various sources, they are gathered in raw form, which makes analysis impossible.

7.2 Rule Mining

The uploaded data set is subjected to the Apriori algorithm. The outcome of this algorithm shows how fatality rates relate to other characteristics like accident type, weather, road surface, lighting, and drunk driving, i.e., under what circumstances the fatality rate is high.

7.3 K Means clustering

One of the most straightforward unsupervised learning techniques to handle the well-known clustering problem is K-means. The process uses a predetermined number of clusters (let's

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

assume k clusters) fixed a priori to categories a given data set. To define k centroids, one for each cluster, is the main notion. These centroids should be positioned deftly because different locations yield various effects. The preferable option is to situate them as far apart from one another as you can. Next, each point from a particular data collection is taken and connected to the closest centroid. The first phase is finished, and an early grouping is carried out when there are no points still open. Now that the clusters produced by the previous phase have their bary centres, we must recalculate k new centroids.

7.4 Naive Bayes Classification

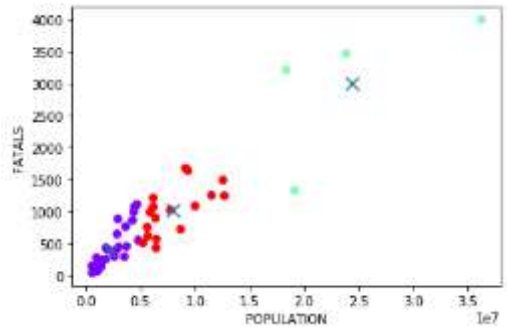
Both a supervised learning method and a statistical classification method are represented by the Bayesian Classification. assumes an underlying probabilistic model, and by calculating the likelihoods of the possible outcomes, it enables us to convey uncertainty about the model in a systematic manner. It can address diagnostic and prognostic issues. With the use of Bayesian classification, one can integrate observable data with previous knowledge and useful learning techniques. The Bayesian Classification approach offers a helpful framework for comprehending and assessing a variety of learning algorithms. It generates precise probability for hypotheses and is resistant to input data noise.

7.5 Visualization the results

Evaluating the outcomes also entails examining the outcomes of the algorithms that were used to process the data set. We can examine the findings from all of this research and offer driving advice to help individuals drive more carefully in areas with a high fatality rate.

8. RESULT

The outcomes of our analysis include frequent itemsets, state clustering in the USA based on fatality rates determined by population and the number of fatal accidents in each state, classification of regions as having a high or low risk of fatal accidents, and a classification label for prediction of unknown data.



9.CONCLUSION

According to statistics, association rule mining, and classification, environmental factors like the state of the road, the weather, and the time of day do not significantly affect the fatality rate, however human factors like whether or not a driver is intoxicated and the sort of collision do.According to the clustering results, some states and regions have a greater fatality rate than others. When driving in certain unsafe areas or places, we might pay closer attention. Through the assignment completed, we came to the conclusion that there is never quite enough facts to support a solid choice. If more data were accessible, such as non-fatal accident data, weather data, distance data, etc., more tests could be conducted and more suggestions could be drawn from the data.

REFERENCES

1. Mussone, L., A. Ferrari, et al. ”An analysis of urban collisions using an artifificial intelligence model”.Accident Analysis and Prevention 31:705-718.
2. Ossenbruggen, P. J., J. Pendharkar, et al. ”Roadway safety in rural and small urbanized areas.”Accidents Analysis and Prevention 33(4):485-498.
3. Sohn, S. and S. Hyungwon. ”Pattern recognition for a road traffic accident severity in Korea”. Ergonomics 44(1): 101-117.
4. Sohn, S. and S. Lee ”Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea...” . Safety Science 41(1): 1-14.
5. Ng, K. S., W. T. Hung, et al. ”An algorithm for assessing the risk of traffic accidents.” Journalof Safety Research 33: 387-410.
6. Eric M Ossiander and Peter Cummings. Freeway speed limits and traffic fatalities in Washington state. Accident Analysis & Prevention, 34(1):13–18, 2002
7. Chang, L. and W. Chen (2005). ”Data mining of tree-based models to analyse freeway accident frequency”. Journal of Safety Research 36:365-375
8. Beshah, T. Application of data mining technology to support RTA severity analysis at Addis Ababa traffic office. Addis Ababa, Addis Ababa University.
9. Chang, L. and H. Wang ”Analysis of traffic injury severity: An application of non-parametric classification tree techniques Accident analysis and prevention ”. Accident analysis and prevention 38(5): 1019-1027.